**AFRL-IF-RS-TR-2004-73**
**Final Technical Report**
**March 2004**


# DNA CODE VALIDATION USING EXPERIMENTAL FLUORESCENCE MEASUREMENTS AND THERMODYNAMIC CALCULATIONS

**The Research Foundation of State University of New York**

**Sponsored by**
**Defense Advanced Research Projects Agency**
**DARPA Order No. J058/30**

*APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.*



**AIR FORCE RESEARCH LABORATORY**
**INFORMATION DIRECTORATE**
**ROME RESEARCH SITE**
**ROME, NEW YORK**

**STINFO FINAL REPORT**

This report has been reviewed by the Air Force Research Laboratory, Information Directorate, Public Affairs Office (IFOIPA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

AFRL-IF-RS-TR-2004-73 has been reviewed and is approved for publication.

APPROVED: /s/
THOMAS RENZ
Project Engineer

FOR THE DIRECTOR: /s/
JAMES A. COLLINS, Acting Chief
Information Technology Division
Information Directorate

# REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No. 074-0188*



| 1. AGENCY USE ONLY (Leave blank) | 2. REPORT DATE | 3. REPORT TYPE AND DATES COVERED |
|---|---|---|
| | MARCH 2004 | **FINAL** **May 03 – Oct 03** |

**4. TITLE AND SUBTITLE**
DNA CODE VALIDATION USING EXPERIMENTAL FLUORESCENCE MEASUREMENTS AND THERMODYNAMIC CALCULATIONS

**5. FUNDING NUMBERS**
G - F30602-03-2-0086
PE - 61101E
PR - DNAI
TA - SB
WU - 03

**6. AUTHOR(S)**
Wendy Pogozelski
Anthony Macula

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
The Research Foundation of State University of New York
35 State Street
Albany NY 12207

**8. PERFORMING ORGANIZATION REPORT NUMBER**
N/A

**9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)**
Defense Advanced Research Projects Agency
3701 North Fairfax Drive
Arlington VA 22203-1714

AFRL/IFTC
26 Electronic Parkway
Rome NY 13441-4514

**10. SPONSORING / MONITORING AGENCY REPORT NUMBER**
AFRL-IF-RS-TR-2004-73

**11. SUPPLEMENTARY NOTES**
AFRL Project Engineer: Thomas Renz/IFTC/(315) 330-3423 Thomas.Renz@rl.af.mil

**12a. DISTRIBUTION / AVAILABILITY STATEMENT**
*APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.*

**12b. DISTRIBUTION CODE**

**13. ABSTRACT** ***(Maximum 200 Words)***
Initial stages were completed in the development of an enabling technology for DNA computing. The technology focuses on construction of a biomolecular architecture to employ new algorithmic paradigms based on DNA hybridization. The intent is to develop a computing basis so that many discrete math problems can be solved in linear real time.

Methods were developed to generate and screen collections of single-stranded DNA sequences called a DNA (n,d) code. Strands of 16 nucleotides were designed such that a code strand would hybridize only with its reverse-complement and would not cross-hybridize with any other strand in the set. Strands were tested for their potential to mispair by measuring fluorescence for every possible pair of strands over varying temperatures in the presence of SYBR Green I, a dye whose fluorescence increases exponentially when bound to double-stranded DNA. Strands with potential to bind to themselves or to another sequence in the set were identified and removed from the DNA (n,d) code. Sets of up to six pooled sequences were also tested to show that this method works with sets of multiple strands. In addition, the thermodynamic parameters of binding were examined using the program PairFold. Free energies of binding are reported.

**14. SUBJECT TERMS**
Biomolecular Computing, DNA Computer, DNA Hybridization

**15. NUMBER OF PAGES** 18

**16. PRICE CODE**

| 17. SECURITY CLASSIFICATION OF REPORT | 18. SECURITY CLASSIFICATION OF THIS PAGE | 19. SECURITY CLASSIFICATION OF ABSTRACT | 20. LIMITATION OF ABSTRACT |
|---|---|---|---|
| UNCLASSIFIED | UNCLASSIFIED | UNCLASSIFIED | UL |

**TABLE of CONTENTS**

## LIST OF FIGURES



## LIST OF TABLES

**SUMMARY**

A DNA code is a collection of single-stranded DNA molecules. In DNA hybridization assays, the formation of any Watson-Crick duplex must be much more energetically favorable than all other possible cross-hybridized duplexes. A DNA code with this property is said to have high binding specificity. In this research, a collection of 16-mer oligonucleotides, each synthesized according to computer-generated blueprints, was tested to validate the code. From this group, a subcollection of DNA oligonucleotides with high binding specificity was extracted.

To determine the affinity of each DNA strand for every other DNA strand in the code, every possible pair of single-stranded DNA molecules was made and combined with SYBR Green I, a dye whose fluorescence increases greatly when bound to double-stranded DNA. These solutions of DNA strands and SYBR Greene were then subjected to conditions designed to denature DNA strands; that is, the thermal agitation was increased such that any helical regions would separate. During the temperature change, fluorescence was monitored using a real-time Polymerase Chain Reaction (PCR) system equipped with a light source, heating and cooling source, fluorescence detector and software. The advantage of this instrument (a Sequence Detection System normally used for real-time PCR), is that it has a 96-well format, allowing for rapid screening. Typically, fluorescence for double-stranded DNA was high and would then decrease as the temperature was lowered. The software created negative derivative plots of this fluorescence. The peak of the resultant curves was the melting temperature. The melting temperature ($T_m$) is defined as the temperature at which DNA is 50% single-stranded and 50% double-stranded. $T_m$ is a useful parameter in thermodynamic calculations and provides an indication of the stability of a helix. Strands with low affinity for one another; that is, non-complements with little tendency to cross-hybridize, had low $T_m$ 's while perfect complements had much higher $T_m$ 's.

These experiments were useful in their ability to identify DNA strands in the original code whose potential to cross-hybridize was too great to be useful in a DNA code. Several sequences were eliminated, resulting in a set of well-behaved strands likely suitable for DNA computing

To complement the experimental work, the DNA sequences were evaluated according to thermodynamic parameters that can be determined from established algorithms using the "nearest neighbor" approach. Using the program PairFold, free energies ($\Delta G$'s) of hybridization were calculated. This approach was useful in identifying strands for which the $\Delta G$ for a cross-hybridized pair was less than four times the $\Delta G$ for a perfect complement pair. The cross-hybridized strands were eliminated from the code.

The two approaches are useful in screening DNA strands for development of architectures for DNA computing.

## INTRODUCTION

Single strands of DNA are, abstractly, (A, C, G, T)-quaternary sequences, with the four letters denoting the respective bases that determine the identity of the nucleotide. DNA sequences are specifically oriented. That is, 5'-AACG-3' is distinct from 5'-GCAA-3. The orientation of a DNA strand is usually indicated by the 5'-3' notation that reflects the asymmetric covalent linking between consecutive bases in the DNA strand. In this paper, when we write DNA molecules without indicating the direction, it is assumed that the direction is 5'→3'. DNA is generally double stranded. That is, each sequence normally occurs with its reverse complement, with reversal denoting that two strands are oppositely directed, and with complementarity denoting that the allowed pairings of letters, opposing one another on the two strands, are{A, T} or {C, G}]. These two combinations represent the canonical Watson-Crick pairings. To obtain the reverse complement of a strand of DNA, one must first reverse the order of the letters and then substitute each letter with its complement. If X is a DNA sequence, we let WC(X) denote its reverse complement. For example, the reverse complement of X=CTATTGAT is WC(X)=ATCAATAG. A Watson-Crick (WC) duplex results from joining reverse complement strands in opposite orientations.

5'-CTATTGAT-3'
3'-GATAACTA-5'

**Figure 1**. Canonical Watson-Crick (WC) duplex

Whenever any two (not necessarily complementary) oppositely directed DNA strands "mirror" one another sufficiently, they are usually capable of coalescing into a DNA duplex. The process of forming a DNA duplex from single strands is referred to as DNA hybridization. The greatest energy of duplex formation is obtained when the two sequences are reverse complements of one another and the DNA duplex formed is a WC duplex. There are however, many instances when the formation of a non-WC duplex is energetically favorable. In this paper, a non-WC duplex is referred to as a cross-hybridized (CH) duplex.

One of the challenges in developing a hybridization-based working architecture for DNA computing is the fact that DNA strands do form CH duplexes. In DNA computing assays, the formation of any WC duplex must be much more energetically favorable than all possible CH duplexes. Examples of CH are shown in Figure 2. Avoidance of these mispairings is crucial for the accuracy of hybridization-based computing methods.

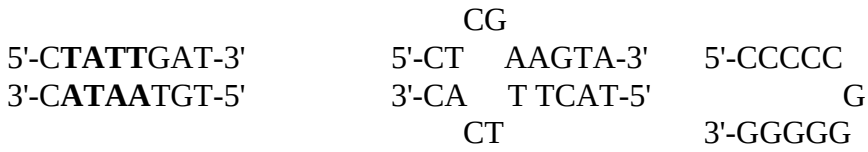


**Figure 2**. Examples of cross-hybridized (CH) duplexes. Left: Mispaired strands (CH duplexes) due to a common subsequence. The hybridized subsequence is shown in bold. Middle: Helix in which a mismatch creates a destabilizing bulge. Right: Classic "hairpin" structure in which one strand folds and pairs with itself.

The first step in creating a well-behaved system is at the level of design - to use the best algorithms to generate the best possible DNA codes. Using software designed by A.Macula and V. Rykov, (Macula, 2003), a set of 13 pairs, (X, WC(X)), of Watson-Crick reverse complementary DNA sequences, called a DNA (16, 6) code was generated. These 26 single stranded sequences (13 pairs) were designed such that no strand had four consecutive Gs and the maximum number base pair bonds in any of the 288 potential CH duplexes[1] was nine.

Screening these strands (or oligonucleotides, as they are called in biochemistry) for actual CH duplexes is a second step in creating DNA molecules that will yield accurate computing results. While many methods exist to test the pairing of DNA (UV-Vis absorbance, etc.), many of these assays are time-consuming and are impractical for testing a large array of DNA sequences. Therefore, we used a fast and highly automated method employing the dye SYBR Green and a Sequence Detection System. Sequence Detection instruments, originally designed for real-time PCR, contain a light source, various filters, a 96-well platform, a somewhat programmable heating and cooling apparatus, and a fluorescence detector capable of monitoring seven absorption and emission wavelengths. One of these wavelengths is that of SYBR Green I, a DNA-binding dye whose fluorescence emission at 510-520 nm increases markedly in the presence of double-stranded DNA. This asymmetrical cyanine dye is an ultrasensitive stain for double-stranded DNA following electrophoresis and is also used to quantify DNA during real-time PCR. Although proprietary (Molecular Probes), SYBR Green I is reported to behave similarly to BEBO (4-[3-methyl-6(bezothiazol-2-yl)-2,3-dihydro(benzo-1,3-thiazole)-2-methylidene)]-1-methyl-1pyridiunium iodide (Bengtsson et al. 2003, structure shown in Figure 3), another cyanine dye, and to have the exact molar mass of Molecular Probes cyanine dye 937 (Zipper et al., 2002; Yue et al., 1997). Our goal was to determine if SYBR Green I fluorescence would enable us to determine the degree to which the DNA strands of our DNA (n, d) code would bind to one another.

[1] There are $\binom{26}{2} - 13 = \frac{26 \cdot 25}{2} - 13 = 312$ possible CH duplexes that consist of distinct strands and there are 26 CH duplexes that consist of the two copies of the same strand.

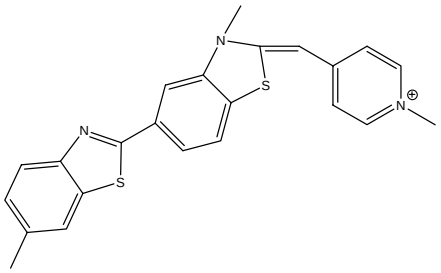

**Figure 3**. Structure of BEBO

Another method for examining the potential for DNA strands to hybridize is to use computer algorithms that examine the thermodynamics of binding. While these methods are commonly used to study strands binding to their perfect reverse-complements, they are less developed in their ability to predict cross-hybridization. However, recent modifications to these programs allow determination of thermodynamic parameters that describe DNA hybridization in quantitative terms.

DNA helix stability is dependent on several factors. The greatest contribution, according to both mathematical models and empirical verification, is the vertical stacking (mainly $\pi-\pi$ interactions) of adjacent base pairs (Borer et al., 1974). Therefore, the identities of the nearest-neighbor bases are crucially important, as they determine this stacking (Freier et al., 1986). The nearest-neighbor model has been extended for heteroduplex stability to include parameters for the interactions that arise with mismatches (Allawi and Santa Lucia, 1997; McDowell and Turner, 1996). We used these models in calculating the free energies of binding for several combinations of sequences.

## METHODS

Sequences for a set of 13 DNA strands and 13 reverse complements were generated using A. Macula's computing methods. The set is called the DNA (16, 6) code and consists of single-stranded oligonucleotides of 16 bases each (16-mers). This list is shown in Table I. These single-stranded sequences were designed such that no strand had four consecutive Gs and the maximum number base pair bonds in any of the 288 potential CH duplexes was nine.

The DNA oligonucleotides were synthesized using phosphoramidite chemistry (InVitrogen). Lyophilized oligonucleotides were dissolved in 10 mM Tris buffer/1 mM EDTA for a concentration of 1 $\mu$g/$\mu$L (0.48 M). All water used in dilutions and buffers was distilled and deionized via a Millipore purification system.

Fluorescent measurements were made on an Applied Biosystems Model 7000 Sequence Detection System. Every combination of the 26 DNA strands was pipetted into the wells of a 96-well optical plate. SYBR Green I is supplied as a concentrated stock with no molecular weight or molar concentration data provided. Efforts to obtain this information from Applied Biosystems were unsuccessful. The optimum amount of

SYBR Green I was, therefore, determined empirically. Each well consisted of 0.5 μg of each oligonucleotide, 1X SYBR Green I Master Mix (Applied Biosystems), and enough distilled deioinized water for a 50 μL volume. The Master Mix includes a passive reference for standardization of the fluorescence. It was important to keep the concentration of SYBR Green constant, as excess SYBR Green can quench the fluorescence signal (Lipsky et al., 2001).

**Table I**. DNA Code (16, 6) with no GGGG or CCCC substring

| X | WC(X) |
|---|---|
| 1. AAAAAAAAAAAAAAAA | C1. TTTTTTTTTTTTTTTT |
| 2. AAAATTTTTTTTAAAA | C2. TTTTAAAAAAAATTTT |
| 3. CGGGAACTTTTTTGGG | C3. CCCAAAAAAGTTCCCG |
| 4. AGGGTCCCTGGTAAAA | C4. TTTTACCAGGGACCCT |
| 5. ATTCCAAAAACCTTAA | C5. TTAAGGTTTTTGGAAT |
| 6. CGGAAACCTAAACGCA | C6. TGCGTTTAGGTTTCCG |
| 7. AACCGTTCAGTCCACA | C7. TGTGGACTGAACGGAA |
| 8. CGCGGGCCCACCAATT | C8. AATTGGTGGGCCCGCG |
| 9. CCTAAAGTTGAAAAAC | C9. GTTTTTCAACTTTAGG |
| 10. CCACTAGTCCGTTTCT | C10. AGAAACGGACTAGTGG |
| 11. CAGGTATAGCAGATTA | C11. TAATCTGCTATACCTG |
| 12. TCCTCGCTGGCATGTC | C12. GACATGCCAGCGAGGA |
| 13. ACTTTTGAGTTGCTAT | C13. ATAGCAACTCAAAAGT |

Fluorescence emission was monitored at 520 nm using the instrument detector and software over a 35 ° temperature window. Measurements were made by slowly increasing the temperature to 60-70 °C over a period of several minutes. The software converted raw fluorescence data (relative to the passive reference) into melting curves by plotting the negative derivative for fluorescence vs. temperature (-dF/dT vs. T). Data were exported to Microsoft Excel for additional analysis. The maximum of each derivative curve corresponds to the melting temperature ($T_m$) of the duplex. The $T_m$ is defined as the temperature at which the DNA is 50% double-stranded.

In addition to testing every possible pair of strands, a pooling experiment was conducted in which up to five different DNA strand populations were mixed. Five wells were used and each well contained an additional DNA sequence.

Several additional experimental parameters were explored. For example, the effect of $MgCl_2$ was tested, since $Mg^{2+}$ ions are known to facilitate the binding of DNA oligonucleotides. We also tested other dyes such as PicoGreen (Molecular Probes) and an alternate vendor preparation of SYBR Green (Sigma). These dyes were dissolved in 1% DMSO.

To truly understand hybridization of DNA strands, one needs to know the thermodynamics of this process. The program PairFold from RNAsoft was used to predict the free energy (ΔG) of binding.

## RESULTS and DISCUSSION

### *Use of Fluorescence to Monitor Hybridization*

SYBR Green I exhibits greatly increased fluorescence when bound to double-stranded DNA. A typical plot of the change in fluorescence as a function of temperature is shown in Figure 4. The sequence $G_{16}$ is being tested for hybridization with $C_{16}$ as well as against the other 24 sequences of the set. Each curve represents a different pair combination of sequences. The magnitude of the change in fluorescence for $G_{16}$ binding to $C_{16}$ is far greater than that of any other combination. Moreover, the temperature which corresponds to the maximum in the curve represents the melting temperature of $T_m$ for this duplex. The $T_m$ is defined as the temperature at which the DNA is 50% single-stranded and 50% double-stranded. This parameter correlates with the thermal stability of the duplex.

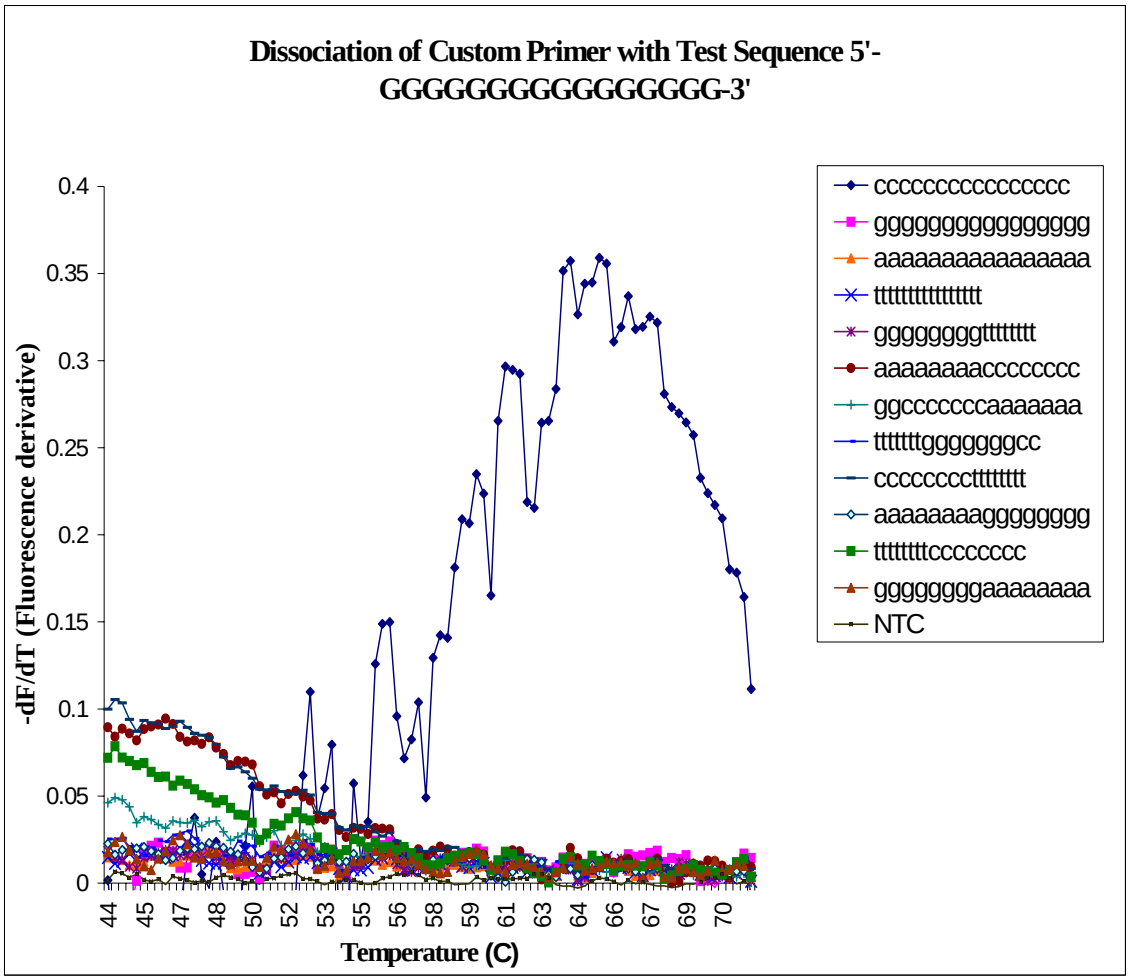


**Figure 4**. Rate of change of fluorescence as a function of temperature for the sequence $G`_{16}$ with its perfect complement $C_{16}$ (large curve) as well as $G_{16}$ paired with every other sequence in the set (smaller curves). The maximum corresponds to the $T_m$.

An example of a curve for a pair combination in which the sequence being tested cross-hybridizes with other strands in the set or pairs to itself is shown in Figure 5. In this plot, the perfect duplex has the highest magnitude for the change in fluorescence, but other duplexes are clearly forming, with measurable $T_m$ values.

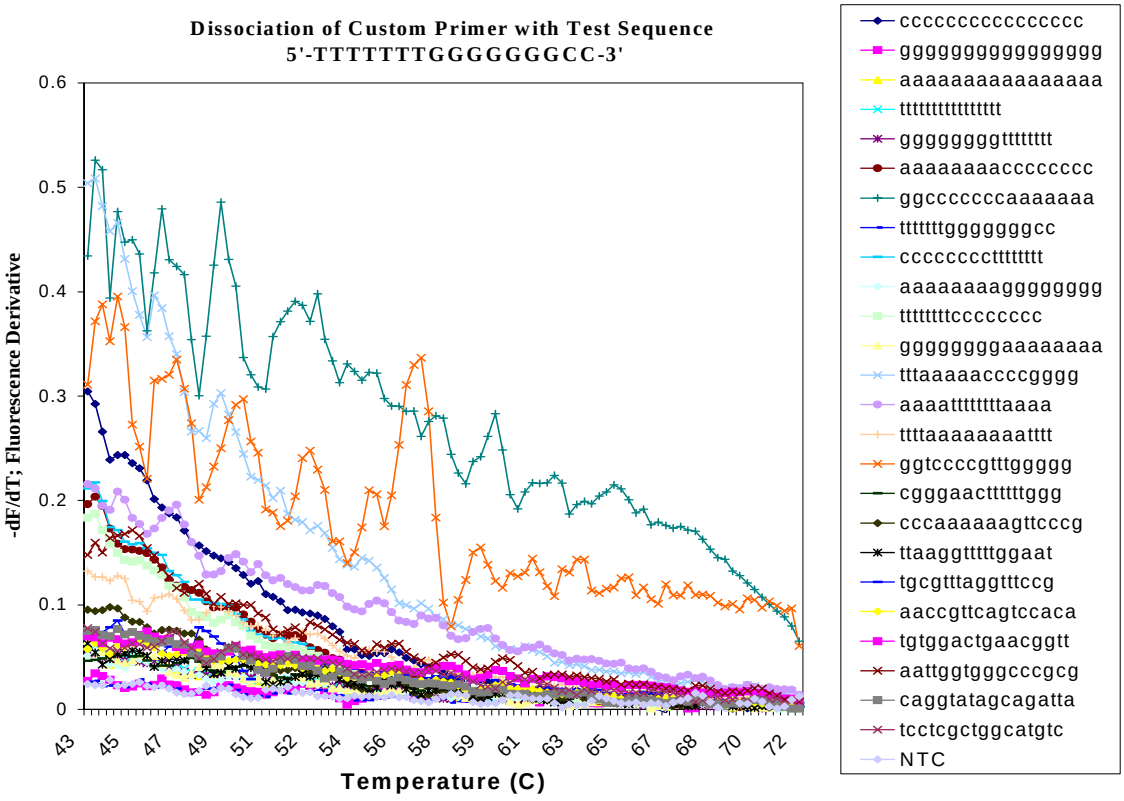


**Figure 5**. Rate of change of fluorescence vs. temperature for the sequence 5'-$T_7G_7C_2$ with its perfect reverse complement 3'-$A_7C_7G_2$ as well as every other pair combination in the set. The highest curve represents the melting curve of the perfect complement.

Sequences X in which the derivative data indicated a CH duplex containing X that was more than 10% of that for the WC duplex containing X were omitted from the DNA(n, d) code. If X was deleted, then WC(X) was also deleted. The resulting collection of nine complementary pairs have the property that each of the nine WC duplexes is at least 10 times as favorable as any of the 162 potential CH duplexes. The revised set of sequences is shown in Table II.

**Table II.** Revised Code of Sequences. Codewords in parenthesis were deleted.

| X | WC(X) |
|---|---|
| (1. AAAAAAAAAAAAAAAA) | (C1. TTTTTTTTTTTTTTTT) |
| (2. AAAATTTTTTTAAAA) | (C2. TTTTAAAAAAAATTTT) |
| 3. CGGGAACTTTTTTGGG | C3. CCCAAAAAAGTTCCCG |
| (4. AGGGTCCCTGGTAAAA) | (C4. TTTTACCAGGGACCCT) |
| 5. ATTCCAAAAACCTTAA | C5. TTAAGGTTTTTGGAAT |
| 6. CGGAAACCTAAACGCA | C6. TGCGTTTAGGTTTCCG |
| 7. AACCGTTCAGTCCACA | C7. TGTGGACTGAACGGAA |
| (8. CGCGGGCCCACCAATT) | (C8. AATTGGTGGGCCCGCG) |
| 9. CCTAAAGTTGAAAAAC | C9. GTTTTTCAACTTTAGG |
| 10. CCACTAGTCCGTTTCT | C10. AGAAACGGACTAGTGG |
| 11. CAGGTATAGCAGATTA | C11. TAATCTGCTATACCTG |
| 12. TCCTCGCTGGCATGTC | C12. GACATGCCAGCGAGGA |
| 13. ACTTTTGAGTTGCTAT | C13. ATAGCAACTCAAAAGT |

In addition to tests of every possible pair of sequences, a pooling experiment was conducted in which populations of up to six different DNA strands were mixed. Figure 6 shows one set of results. In this melting curve, the data that is flat on the bottom of the graph show derivative fluorescence for combinations of two, three, four and five non-complementary strands. The non-zero curve represents derivative fluorescence for a combination of six sequences, two of which are reverse complements of one another. Several interesting observations can be made from these data. First, the robustness of the SYBR Green method extends to conditions in which multiple DNA sequences are pooled. Secondly, the fact that the fluorescence derivative of the non-complements is so small relative to that of the perfect complements demonstrates that when a perfect complement is present in a solution, the probability of mispairing to another strand is greatly reduced. This fact has enormous implications for DNA computing.

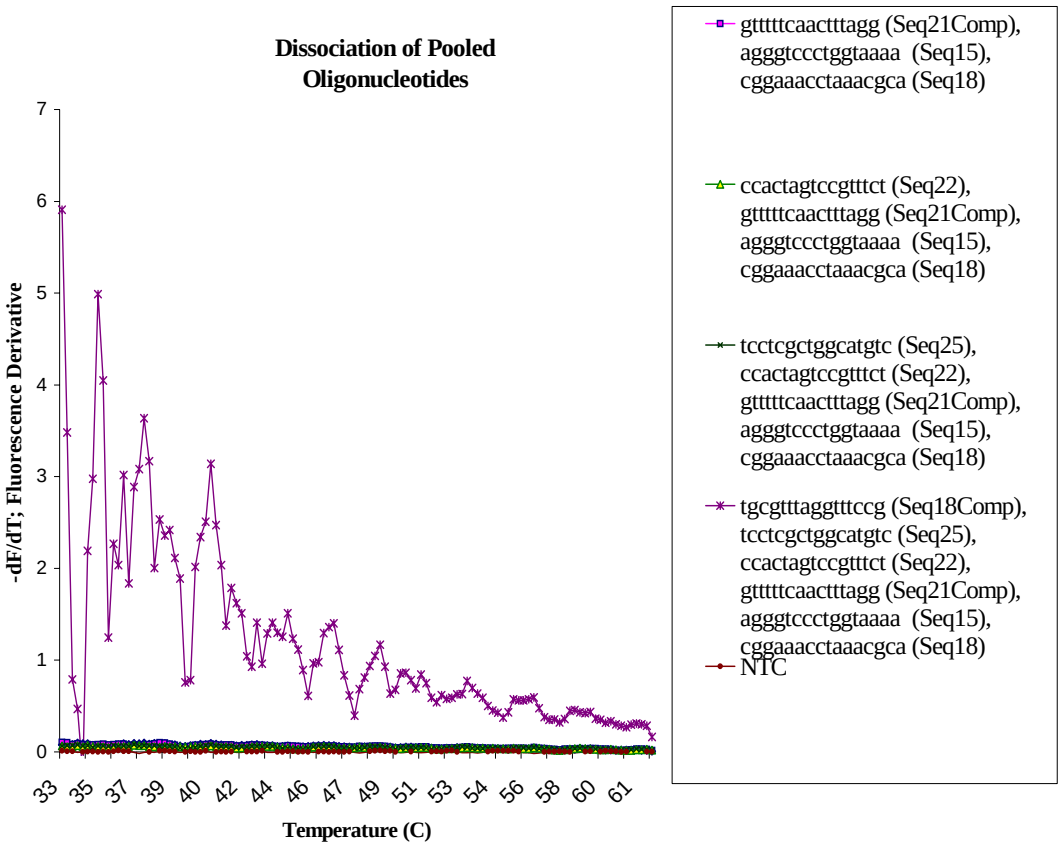


**Figure 6**. Pooling Experiment. DNA strands were mixed as pairs, then in groups of three, four, five, and six populations of strands. The six strands included the reverse complement of one of the other strands.

Several additional experimental parameters were explored to try to optimize these results. For example, the effect of $MgCl_2$ was tested, since $Mg^{2+}$ ions are known to facilitate the binding of DNA oligonucleotides. However, divalent ions can stabilize imperfect hybrids as well. We found that addition of 1.5 mM $MgCl_2$ had an adverse effect on results, so future experiments did not include additional magnesium. Various temperature ranges were also tested. It was found that the melting temperatures of unstable duplexes (mismatches) required very low temperature measurements.

We also tested other dyes such as PicoGreen (Molecular Probes) and an alternate preparation of SYBR Green (Sigma), since the SYBR Green from Applied Biosystems is expensive and comes in a kit with other components that are not useful for this work. These experiments did not yield useful data, perhaps because it was difficult to approximate the concentration of the Applied Biosystems SYBR Green as well as the components of the buffer.

The use of SYBR Green, along with the Sequence Detection System to predict mispairing, is fast, easy and the method is robust. The dye is stable at temperatures needed for denaturation experiments and the dye is readily available. The use of the Sequence Detection System also allows for the processing of large numbers of samples. Disadvantages include the expense of the dye and the fact that because the dye is proprietary, no information about structure, binding details and concentration are yet available. These data can be determined experimentally, however and we intend to investigate the number of SYBR Green molecules that can bind per turn of helix, the mode of binding and the sequence dependence of binding. Future experiments will probe the concentration dependence of SYBR Green I. Other cyanine dyes such as BEBO aggregate at high concentrations (Bengtsson et al., 2003), although the aggregation can be somewhat alleviated by DMSO. We intend to find the limit of solubility of SYBR Green in this system and determine the optimum conditions for future experiments.

These measurements are made as the temperature is increased from approximately 25-35 °C to 60-70 °C. This particular instrument does not permit altering the heating program to collect data in the direction of decreasing temperature. There are other real-time PCR machines that can accomplish this task, but our system does not allow it. Hence, we have not tested whether any hysteresis occurs with this method. Hysteresis is generally not found with short DNA oligonucleotides like the 16-mers used here; however, future plans include using a fluorescence spectrometer to testing SYBR Green binding under different annealing and denaturation conditions for possible hysteresis. We also intend to use this latter method to assess the error in SYBR Green fluorescence.

***PairFold Theoretical Calculations***

To examine the thermodynamics of strand hybridization for this code, the program PairFold from the RNAsoft suite of programs was used. The program can be found at http://www.RNAsoft.ca. PairFold predicts the minimum free energy secondary structure of two input DNA strand sequences and can be used to predict interactions between the strands (Andronescu et al., 2003). It is based on the free energy model (Zucker et al., 1999), predicting that under fixed conditions of temperature and ionic strength, two DNA strands will pair to a structure that minimizes the free energy. This free energy is determined from the sum of the energies of stacked pairs. The advantage of PairFold is that it takes two sequences as input, as opposed to other programs which handle only one sequence. The algorithm is based on the Zuker-Steiger algorithm for single molecules of RNA (Zucker and Steiger, 1981). Calculations were performed for all pair combinations of strands. Results for the revised code (from Table II) are shown in Table III (next page). The $\Delta G$ values for the perfect complements are generally at least four times higher than those of sequence pairs that are not expected achieve stable hybridization. The small $\Delta G$ values for these non-complementary sequences show that in the absence of a strong complement, DNA strands will show limited affinity for one another.

These calculations were made using the default salt concentration of 1 M. We would like to repeat these calculations using an ionic strength value determined from experiments with SYBR Green. Because Applied Biosystems refuses to give us information regarding ionic strength in its SYBR Green preparation and buffer, we will need to

determine this value empirically. Our thermodynamic calculations and experimental results can then be more meaningfully compared.

In addition, we have also begun to explore the use of a new program called MeltCalc from which we can estimate the parameters $T_m$, $\Delta G$, $\Delta H$, and $\Delta S$. These parameters will also be useful in establishing a method to quantitatively predict the potential for DNA strands to hybridize in future DNA computing architectures.

**Table III. Results of PairFold Calculations**
**ΔG Values (kcal/mol) at 35 °C**

| Strand | 3 | 5 | 6 | 7 | 9 | 10 | 11 | 12 | 13 | C3 | C5 | C6 | C7 | C9 | C10 | C11 | C12 | C13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **3** | -1.1 | -6.4 | -2.8 | -4.0 | -5.8 | -3.4 | -1.9 | -2.5 | -4.4 | **-19.0** | -2.9 | -5.2 | -2.1 | -1.9 | -2.7 | -2.8 | -2.9 | -4.9 |
| **5** | | -0.6 | -3.9 | -1.8 | -2.3 | -2.8 | -4.2 | -2.2 | -4.0 | -2.1 | **-15.6** | -6.2 | -3.6 | -4.4 | -2.6 | -0.9 | -3.9 | -2.1 |
| **6** | | | -2.6 | -4.3 | -2.3 | -5.3 | -2.6 | -2.8 | -3.1 | -3.5 | -6.4 | **-19.8** | -3.6 | -4.4 | -2.8 | -2.6 | -5.7 | -2.8 |
| **7** | | | | -2.0 | -1.9 | -2.2 | -2.9 | -2.9 | -1.8 | -1.6 | -3.3 | -3.2 | **-19.8** | -2.1 | -7.0 | -2.5 | -3.7 | -2 |
| **9** | | | | | -2.2 | -2.5 | -2.5 | -1.4 | -4.0 | -2.0 | -5.1 | -3.6 | -2.6 | **-15.7** | -2.6 | -2 | -3.3 | -5.5 |
| **10** | | | | | | -5.2 | -2.1 | -1.9 | -2.5 | -2.4 | -2.4 | -2.2 | -6.3 | -2.4 | **-18.6** | -1.5 | -3.7 | -2.3 |
| **11** | | | | | | | -5.4 | -3.9 | -5.2 | -2.2 | -1.6 | -2.3 | -4.3 | -1.9 | -2.1 | **-17** | -3.2 | -3.1 |
| **12** | | | | | | | | -5.0 | -2.9 | -2.2 | -2.7 | -4.5 | -2.7 | -2.6 | -2.7 | -2.8 | **-22** | -3.9 |
| **13** | | | | | | | | | -2.5 | -5.0 | -2.7 | -2.1 | -2.7 | -4.9 | -2.7 | -2.5 | -3.1 | **-17** |
| **C3** | | | | | | | | | | -1.6 | -6.7 | -2.2 | -3.9 | -5.3 | -4.5 | -1.3 | -2.9 | -3 |
| **C5** | | | | | | | | | | | -0.5 | -2.7 | -1.9 | -2.9 | -2.9 | -4.5 | -3 | -3.6 |
| **C6** | | | | | | | | | | | | -2.4 | -5.7 | -2.2 | -5.9 | -4.7 | -2.9 | -2.4 |
| **C7** | | | | | | | | | | | | | -4.3 | -1.9 | -3.6 | -3.1 | -4 | -2.3 |
| **C9** | | | | | | | | | | | | | | -1.7 | -2.8 | -3.3 | -2.3 | -4.2 |
| **C10** | | | | | | | | | | | | | | | -6.2 | -2.6 | -3.3 | -2.6 |
| **C11** | | | | | | | | | | | | | | | | -2.5 | -4.1 | -5 |
| **C12** | | | | | | | | | | | | | | | | | -5 | -3.2 |
| **C13** | | | | | | | | | | | | | | | | | | -3.1 |

## CONCLUSIONS

Thirteen pairs of DNA oligonucleotides were synthesized according to a computer-generated set of 13 pairs, (X, WC(X)), of Watson-Crick reverse complementary quaternary sequences over the alphabet {A, C, G, T}. All 338 potential CH duplexes and 13 WC duplexes where tested for their stability by measuring the rate of change of fluorescence in the presence of SYBR Green and by calculating the free energy of hybridization using the nearest-neighbor approach. Based on our experimental results, four oligonucleotide sequences and their complements were deleted because at least one member of the deleted pair appeared in a CH duplex with a stability that was more than 10% of one of the 13 WC duplexes. Thus the remaining set of 9 (X, WC(X)) pairs represent an experimentally verified DNA code with high binding specificity.

These results point to the usefulness of using fluorescence for validating DNA codes for high binding specificity. Our preliminary tests on SYBR Green analysis of pools of several DNA sequences indicate the potential of using SYBR green in high throughput DNA code validation protocols based on mathematical group testing methods.

**REFERENCES**


Allawi, H.T., SantaLucia, J., Jr. (1997) Thermodynamics and NMR of internal G-T mismatches in DNA. Biochemistry, 36, 10581-10594.

Andronescu, M.; Aguirre-Hernández, R.; Condon, A.; Hoose, H.H. (2003) RNAsoft: a suite of RNA secondary structure prediction and design software tools. Nucleic Acids Research, 31, 13, 3416-3422.

Bengtsson, M., Karlsson, H.J., Westman, G. and Kubista, M. (2003) A new minor groove binding asymmetric cyanine reporter dye for real-time PCRNucleic Acids Research, 31(8), e45.

Borer, P.N., Gengler, B., Tinoco, I., Jr. (1974) Stability of ribonucleic acid double-stranded helices. J. Mol. Biol. 86, 843-853.

Freier, S.M., Sugimoto, N., Sinclair, A., Alkema, D., Neilson, T., Kierzek, R. et al. (1986). Stability of XGCGCp, GCGCYp, and XGCGCYp helixes: an empirical estimate of the energetics of hydrogen bonds in nucleic acids. Biochemistry, 25, 3214-3219.

Lipsky, R.H., Mazzanti, C.M., Rudolph, J.G., Xu, K., Vyas, G., Bozak, D., Radel, M.Q., Goldman, D. (2001) DNA melting analysis for detection of single nucleotide polymorphisms, Clinical Chemistry, 47, 635-644.

Macula, A.J. (2003) US Air Force Office of Scientific Research Technical Report, DNA TAT Codes TR-2003-57, AFRL-IF-RS, http://stinet.dtic.mil/cgi-bin/fulcrum_main.pl

McDowell, J.A., Turner, D.H. (1996) Investigation of the structural basis for thermodynamic stabilities of tandem GU mismatches: solution structure of $(rGAGGUCUC)_2$ by two-dimensional NMR and simulated annealing. Biochemistry, 35, 14077-14089.

Yue, S.T., Singer, V.L., Roth, B.L., Mozer, T.J., Millard, P.J., Jones, L.J., Jin, X. and Haugland, R.P. (1997). Substituted unsymmetrical cyanine dyes with selected permeability, US-patent 5658751.

Zipper, H., Lämmle, K., Buta, C., Brunner, H., Bernhagen, J., Vitzthum, F. Investigations on the binding of SYBR Green I to double-stranded DNA, Proceedings, German Society for Biochemistry and Molecular Biology & German Society for

Experimental and clinical Pharmacology and Toxicology, Halle, Germany, Sept. 7-10, 2002.

Zucker, M., Matthews, D.H., Turner, D.H. (1999) Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide. Braciszewski, J. and Clark, B.F.C. (eds) In *RNA Biochemistry and Biotechnology*, NATO ASI Series, Kluwer Academic Publishers.

Zucker, M. and Stiegler, P. (1981) Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acids Res. 9, 133-148.